Identifying an Unknown Static Process Having Uniformly Distributed Observational Errors

H. STALFORD

Radar Analysis Staff
Radar Division

April 2, 1975

NAVAL RESEARCH LABORATORY
Washington, D.C.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NRL Report 7784	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) IDENTIFYING AN UNKNOWN STATIC PROCESS HAVING UNIFORMLY DISTRIBUTED OBSERVATIONAL ERRORS		5. TYPE OF REPORT & PERIOD COVERED An interim report on a continuing NRL problem. 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Harold L. Stalford		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, D.C. 20375		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NRL Problem Number B01-15 RR 014-02-01
11. CONTROLLING OFFICE NAME AND ADDRESS Department of the Navy Office of Naval Research Arlington, VA 22217		12. REPORT DATE April 2, 1975 13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Identification
Unknown process
Black box

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The problem of identifying a black box having uniformly distributed errors in the observations is investigated. In particular, those black boxes are studied for which the space of candidate models is infinite dimensional. Since such black boxes cannot be identified deterministically by using a finite amount of data, the concept of a probability of identification is followed. An adequacy criterion for modeling is defined in which a model is adequate provided it passes a finite sequence of tests, the length of which depends on the desired probability of identification. An algorithm is presented for determining an adequate model within the space of polynomial functions.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

CONTENTS

IDENTIFYING AN UNKNOWN STATIC PROCESS HAVING UNIFORMLY DISTRIBUTED OBSERVATIONAL ERRORS

1. INTRODUCTION

The black box identification problem is that of determining, on the basis of a finite amount of experimental input-output data, an adequate model of the unknown process (the black box). The principal elements of this problem are a space of candidate models, the spaces of admissible inputs and possible outputs, and a criterion for adequate modeling, Zadeh [1,2]. Identification problems can be grouped according to whether the space of candidate models is finite or infinite dimensional.

In the last decade extensive research has been conducted on the finite case, and many remarkable results have been achieved. For instance, process parameter estimation and state estimation fall into this group. They deal with estimating a finite number of unknown scalar quantities of a physical process. In such problems the space of candidate models is representable as a subset of a finite-dimensional Euclidean space. The literature contains a number of excellent survey articles on methods for resolving these problems. For example, Eykhoff, Van der Grinten, Kwakernaak, and Veltman [3] discuss a number of techniques proposed for parameter estimation and cite some examples of industrial models. Eykhoff [4] summarizes some important properties of the parameter-estimation problem and gives a brief discussion on process state estimation. Cuenod and Sage [5] present a survey of some of the computational problems and procedures of process-parameter identification. Balakrishnan and Peterka [6] compare the principal ideas of different approaches to the identification problem. Åström and Eykhoff [7] give an overview of the field of identification as it relates to control-engineering applications. The articles in Eykhoff [8] cover applications in aeronautics, biology, chemistry, economics, engineering and physics. The text of Eykhoff [9] gives an excellent account of the theory of parameter estimation.

Only a modicum of results has been discovered for identification problems in which the space of candidate models is infinite dimensional. In such problems less knowledge about the process is assumed to be known *a priori*. That is, it is not assumed that the process evolves according to a linear or quadratic law or that the process has a known finite number of unknown quantities. Rather, loosely speaking, the black box is less transparent than in the finite case. (From the viewpoint of *a priori* knowledge, it is more appropriate to speak of unknown processes as gray boxes varying in shades rather than as black boxes. Nevertheless, it is conventional to refer to all unknown processes as black boxes.) In the finite case a finite amount of input-output data is sufficient for the purpose of constructing deterministically the identity of a black box. This is not true for the infinite case. For example, necessary and sufficient conditions are given by Stalford and Leitmann [10] for representing a black box as a dynamical system governed by ordinary differential equations. Those conditions cannot be verified by a finite amount of input-output data. The publication by Gold [11] is an investigation on identifying a black box in the limit (i.e., experimenting on the black box indefinitely). The limiting method provides no means of establishing that a constructed model is the identity even though the correct model actually may be obtained after finite experimentation. Although a black box belonging to the infinite case cannot be identified deterministically with finite data, it often can be identified with probability 1 only using finite data. For instance, Stalford and Kullback [12] established by means of an algorithm that noise-free static black boxes representable by polynomials are identifiable with probability 1. The goal of this report is to extend that investigation to static processes having uniformly distributed observational errors.

A precise statement of the problem under investigation is given in the next section. A method of testing a candidate model for its probability of being the identity is presented in Section 3. The value of selecting random input data is discussed in Section 4. An adequacy criterion for modeling is described in Section 5. An algorithm for identifying an adequate polynomial model of an unknown static process is established in Section 6.

2. PROBLEM STATEMENT

For any integer m, we let R^m denote an m-dimensional Euclidean space. Let the integers m_i and m_0 represent respectively the dimensions of the input and output vectors of a black box. Let X be a compact subset of R^{m_i}, and let Y be a Borel-measurable subset of R^{m_0}. The sets X and Y are the sets of all admissible inputs and possible outputs respectively. Let $\mathfrak{L}$ denote the space of all essentially bounded Borel-measurable functions with domain X and range in Y. Throughout this report we let $\mathfrak{F}$ denote a subspace of $\mathfrak{L}$.

Let ϵ be a positive real number. For the purposes of this report, we define a black box B to be a quadruple $\{X, f, Y, \epsilon\}$, where f is an unknown member of $\mathfrak{F}$. The mapping f is the input-output relationship of the black box B, and it is termed the identity of B. The number ϵ represents the absolute bound on uniformly distributed observational errors. Let such errors be denoted by η. This random variable satisfies $|\eta| \leqslant \epsilon$, where $|\eta|$ is the Euclidean norm of η in R^{m_0}.

Let the observation of the output for an input $x \in X$ be denoted by $\Theta(x)$. The observations satisfy the equation

$$\Theta(x) = f(x) + \eta \tag{1}$$

for all inputs $x \in X$.

The space $\mathfrak{L}$ may seem too large a space of functions from which identities f are permitted to belong, particularly when only elementary models of f are desirable. Elementary models are those designed with elementary functions such as polynomials, exponentials, etc. According to the Stone-Weierstrass theorem of analysis, a continuous function can be approximated uniformly by polynomials, with arbitrary accuracy. By

Lusin's theorem of measure theory, a Borel-measurable function is equal to a continuous function except on a subset of X of arbitrarily small measure. Thus for any f contained in $\mathfrak{L}$, we can approximate it arbitrarily close by a polynomial model if desired.

For f, g contained in $\mathfrak{L}$, we make the definition $f = g$ if and only if f and g are identical almost everywhere. Thus if $f \neq g$, then f and g differ on a set of positive measure.

For the black box $B = \{X, f, Y, \epsilon\}$, let P be a prior distribution on the space $\mathcal{F}$ of candidate models. In the sequel, this *a priori* probability is denoted by $P(f = g)$, $g \in \mathcal{F}$.

Problem. *Let* q *denote a real number in* (0, 1). *Given a black box* $\mathrm{B} = \{\mathrm{X}, \mathrm{f}, \mathrm{Y}, \epsilon\}$ *with* $\mathrm{f} \in \mathcal{F}$ *and the a priori distribution* P *over* $\mathcal{F}$, *find, by analyzing only a finite number of observations, a model* $\mathrm{g} \in \mathcal{F}$ *such that* $\mathrm{f} = \mathrm{g}$ *with probability* q.

A diagram of our problem is given in Fig. 1. Randomly chosen inputs will be used in resolving this problem. An asterisk is used as a superscript on an input variable to denote a randomly chosen input. It is tacitly assumed that members of a finite sequence of randomly chosen inputs $\{x_0^*, x_1^*, \ldots, x_n^*\}$ are independently generated. We assume unlimited freedom in selecting the inputs from X.

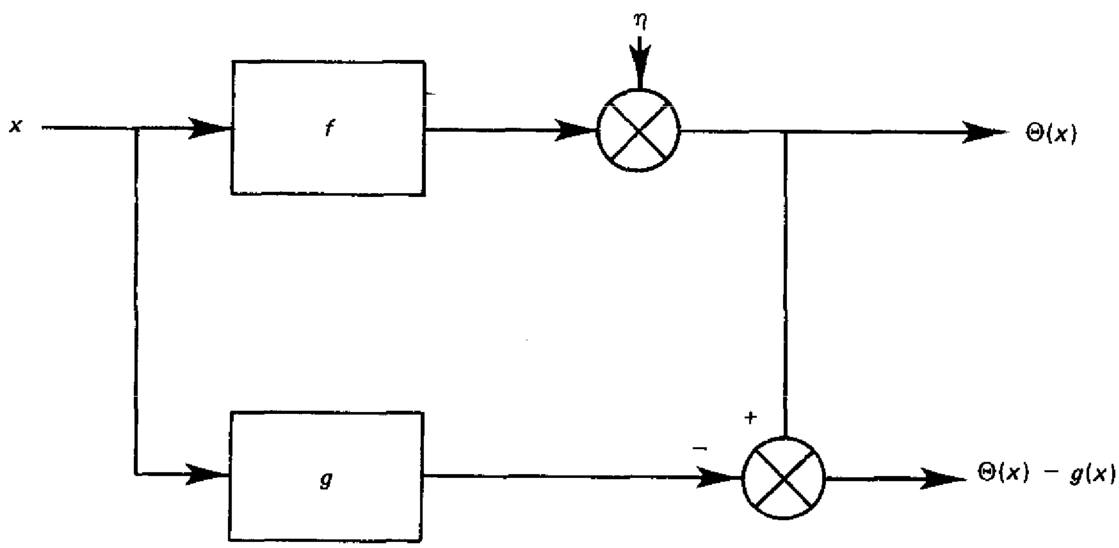

Figure 1

We need the following notation. Let $y_0 \in R^{m_0}$, and let r be a positive real number. The closed ball centered at y_0 with radius r is denoted by

$$B(y_0, r) = \{y \in R^{m_0} : |y - y_0| \leqslant r\}.$$

For a given measurable subset, say C, of R^{m_0}, we let $\nu(C)$ denote the Lebesgue measure of C.

3. TESTING A CANDIDATE FOR THE IDENTITY

Let g be contained in $\mathcal{F}$. We suppose that $g(x)$, $x \in X$, can be observed without error. We seek a means of increasing our confidence in $f = g$ above the *a priori* value $P(f = g)$. Let x^* be an input, and consider the difference $|\Theta(x^*) - g(x^*)|$. If the difference is greater than ϵ, then the model g is clearly not the identity f. On the other hand, if it is less than or equal to ϵ, then our confidence in g being the identity is enhanced. The degree of enhancement is the object of this section.

Let E denote the event that an input x results in the inequality

$$|\Theta(x) - g(x)| \leqslant \epsilon \tag{2}$$

being met. Write

$$A = \{x \in X : |f(x) - g(x)| \leqslant 2\epsilon\}.$$

The event E cannot occur whenever x is not a member of A. For $x \in A$ the event E can occur with the probability

$$P(E|x) = \frac{\nu(B(g(x), \epsilon) \cap B(f(x), \epsilon))}{\nu(B(f(x), \epsilon))}.$$

For the case that $m_0 = 1$, this equation reduces to

$$P(E|x) = 1 - \frac{|f(x) - g(x)|}{2\epsilon}.$$

For a randomly chosen input x^* the event E has the probability

$$P(E) = \int_X P(E|x) P_u(x)\, dx$$

of occurring, where P_u is the probability density of choosing a random input. For a uniform density we have $P_u(x) = 1/\mu(X)$ for all $x \in X$, where $\mu(X)$ is the Lebesgue measure of X. Conditioned on $f \neq g$, the event E occurs with the probability

$$P(E|f \neq g) = \frac{1}{\mu(X)} \int_A \frac{\nu(B(g(x), \epsilon) \cap B(f(x), \epsilon))}{\nu(B(f(x), \epsilon))}\, dx\,.$$

Note that $P(E|f \neq g) < 1$, since $f \neq g$ implies that f and g differ on a set of positive measure.

Theorem 1. *If a random input results in the event* E, *then the probability that* f = g *conditioned on the happening* E *is given by*

$$P(f=g|E) = \frac{P(f=g)}{P(f=g) + [1 - P(f=g)]P(E|f \neq g)}.$$

Proof. The probability that $f = g$ conditioned on E is given by the Bayes' formula as

$$P(f=g|E) = \frac{P(E|f=g)P(f=g)}{P(E|f=g)P(f=g) + P(f \neq g)P(E|f \neq g)}$$

If $f = g$, then $A = X$ and $P(E|x) = 1$ for all $x \in X$; therefore $P(E|f=g) = 1$. The assertion now follows since $P(f \neq g) = 1 - P(f=g)$.

It follows from this theorem that

$$P(f=g|E) > P(f=g)$$

whenever $P(f=g) \neq 0$ or 1. Thus our confidence in $f = g$ is increased by the occurrence of the event E.

If $P(f=g) = 0.5$, then

$$P(f=g|E) = \frac{1}{1 + P(E|f \neq g)}.$$

In the next section we investigate the use of administering randomly chosen inputs over n stages.

4. SELECTING A SEQUENCE OF INPUT DATA

We call a test that operation of administering an input to the black box and to the model, observing the output of each, and checking to see if the event E occurs. We let n represent the length of a given series of tests.

Since $f \neq g$ implies that f and g differ on a set of positive measure, it follows that $P(E|x^*) < 1$ unless $f = g$. Thus there is the positive probability $1 - P(E|x^*)$ that the event E does not occur whenever $f \neq g$. As a consequence, if x^* is used as the test input n consecutive times, then as n increases without bound we are sure to find out that $f \neq g$ when this actually is the case. This remark raises a question. Which method gives the highest probability of $f = g$, conditioned on the event E occurring at each of n tests: the method of administering different randomly chosen inputs for each test or the method of applying the same randomly chosen input for all n tests? We establish in the proof of the next theorem that the former method gives the highest probability.

Let $S_d = \{x_1^*, x_2^*, \ldots, x_n^*\}$ denote a sequence of distinct (independent) randomly chosen inputs. Let $S_s = \{x^*, x^*, \ldots, x^*\}$ denote the sequence of length n of the same randomly chosen input x^*. For $i = 1, 2, \ldots, n$, let $E_i(S_d)$ denote the occurrence of the event E for the input $x_i^* \in S_d$, and let $E_i(S_s)$ denote the occurrence of the event E for the ith application of the input $x^* \in S_s$.

Theorem 2. *For the sequences* S_d *and* S_s *of length* n, *we have*

$$P\left(f = g \,\middle|\, \bigcap_{i=1}^{n} E_i(S_d)\right) = \frac{P(f = g)}{P(f = g) + [1 - P(f = g)]\, P^n(E|f \neq g)} \tag{3}$$

$$P\left(\bigcap_{i=1}^{n} E_i(S_d) \,\middle|\, f \neq g\right) < P\left(\bigcap_{i=1}^{n} E_i(S_s) \,\middle|\, f \neq g\right)$$

$$P\left(f = g \,\middle|\, \bigcap_{i=1}^{n} E_i(S_d)\right) > P\left(f = g \,\middle|\, \bigcap_{i=1}^{n} E_i(S_s)\right).$$

Proof. Since the random inputs of S_d are randomly and independently chosen, it follows that

$$P\left(\bigcap_{i=1}^{n} E_i(S_d) \,\middle|\, f \neq g\right) = \prod_{i=1}^{n} P(E_i(S_d)|f \neq g) = P^n(E|f \neq g).$$

This together with Theorem 1 gives Eq. (3).

The probability of the event

$$\bigcap_{i=1}^{n} E_i(S_s)$$

conditioned on $x^* = x \in A$ is given by

$$P\left(\bigcap_{i=1}^{n} E_i(S_s) \,\middle|\, x^* = x\right) = \left[\frac{\nu(B(g(x), \epsilon) \cap B(f(x), \epsilon))}{\nu(B(f(x), \epsilon))}\right]^n$$

Integrating over all $x \in X$ gives

$$P\left(\bigcap_{i=1}^{n} E_i(S_s) \,\middle|\, f \neq g\right) = \frac{1}{\mu(X)} \int_A \left[\frac{\nu(B(g(x), \epsilon) \cap B(f(x), \epsilon))}{\nu(B(f(x), \epsilon))}\right]^n dx.$$

Applying the Hölder inequality of integration theory establishes that

$$P^n(E|f \neq g) < P\left(\bigcap_{i=1}^{n} E_i(S_s)\,\middle|\, f \neq g\right).$$

This expresses that there is a higher probability of finding out that $f \neq g$ (when this is the case) if distinct random inputs are used rather than if a single random input is used for all tests. It follows from Theorem 1 that

$$P\left(f = g \,\middle|\, \bigcap_{i=1}^{n} E_i(S_s)\right) = \frac{P(f=g)}{P(f=g) + [1 - P(f=g)]\, P\left(\bigcap_{i=1}^{n} E_i(S_s)\,\middle|\, f \neq g\right)}$$

From this equality, together with the previous inequality and Eq. (3), it follows that the last inequality of Theorem 2 is met.

For $P(f = g) \neq 0$ we have the corollary

$$\lim_{n\to\infty} P\left(f = g \,\middle|\, \bigcap_{i=1}^{n} E_i(S_d)\right) = 1\,.$$

5. ADEQUACY CRITERION FOR MODELING

The probability $P(E|f \neq g)$ is the probability that the observed output $\Theta(x^*)$ will be within the error ϵ of the output $g(x^*)$ of the model. For any $x \in X$ the value $g(x)$ is referred to as the predicted output of the black box. In the modeling of unknown systems, it is desirable to have a model with which the observations can be predicted within an error of ϵ. Let P_s denote the probability that the event E occurs whenever some input x^* is applied to the black box. Then $P_s = P(E|f \neq g)$. We call P_s the probability of a successful test.

For n tests we write

$$P_I(n) = P\left(f = g \,\middle|\, \bigcap_{i=1}^{n} E_i(S_d)\right).$$

We call $P_I(n)$ the probability of identification resulting from the occurrence of the event E at each of n consecutive tests.

Our main equation, Eq. (3), can be rewritten as

$$P_I(n) = \frac{P(f=g)}{P(f=g) + [1 - P(f=g)]\, P_s^n}\,. \tag{4}$$

In the modeling of a particular physical process, we cannot in general know with probability 1 that a certain model is the identity of a black box. Rather, we must be content with a model that has a high probability of a successful test and that gives us a high level of confidence in it being a replica of the black box.

For given values of P_s, $P(f = g)$, and $P_I(n)$, Eq. (4) provides a criterion for testing the adequacy of a certain model g. Suppose we desire a given value P_s as the probability of a successful test. Beginning with an initial confidence $P(f = g)$, we would like to increase our confidence to the level $P_I(n)$. The number of tests n necessary to provide the additional confidence is given by

$$n = \frac{\log\left[\dfrac{\dfrac{1}{P(f=g)} - 1}{\dfrac{1}{P_I(n)} - 1}\right]}{\log\left[\dfrac{1}{P_s}\right]}, \tag{5}$$

that is, n is the solution of Eq. (4). For example, with $(P_s, P(f = g), P_I(n)) = (0.99, 0.5, 0.95)$, we have from Eq. (5) that $n = 293$ tests. If $P(f = g)$ is increased to 0.8, then $n = 155$. However, if $P(f = g)$ is decreased to 0.2, then $n = 431$. In view of this we make the following definition.

Definition 1. *A model* g *of a black box is adequate with respect to the values* $(P_s, P(f = g), P_I(n))$ *provided the event* E *occurs at each of* n *tests, where* n *is given by Eq. (5).*

Thus adequacy is a function of the probability of a successful test, an *a priori* confidence in $f = g$, and an *a posteriori* confidence in $f = g$. Note that n is an increasing function of P_s and $P_I(n)$ and is a decreasing function of $P(f = g)$.

6. IDENTIFICATION ALGORITHM

In this section we present an algorithm for identifying a polynomial function of a single variable. Afterwards, the vector case is discussed.

Let $\mathfrak{F}$ denote the class of polynomial functions of one variable. For $g \in \mathfrak{F}$, let $P(f = g)$, a nonzero value, be the initial confidence in the polynomial model $g \in \mathfrak{F}$ being the identity of a given black box (X, f, Y, ϵ), where $f \in \mathfrak{F}$, X is compact in R^1, and $Y = R^1$. Let P_s be a preset level for the probability of a successful test. Let $P_I(n) > P(f = g)$ be the *a posteriori* confidence required of any model g of f. Our goal is to construct an adequate model g, that is, one for which the event E occurs after each of n tests, where n satisfies Eq. (5). Our goal is reached with the algorithm given below in Theorem 3. First, we need some lemmas.

The following lemma establishes that a single Lipschitz constant K is satisfied by all polynomials of degree less than $m + 1$ that satisfy compact constraints at $m + 1$ distinct points.

Lemma 1. *Let* m *be any positive integer. Let* X *be a compact subset of* R. *For* $i = 0, 1, 2, \ldots, m$, *let* $y_i \in X$ *be distinct, and let* I_i *denote a compact interval of* R^1. *Then there exists an integer* K *such that*

$$|p(x) - p(y)| \leq K|x - y|, \; x, y \in X,$$

for all polynomials p *of degree less than* m + 1 *satisfying*

$$p(y_i) \in I_i, \quad i = 0, 1, 2, \ldots, m.$$

Proof. A polynomial p of degree less than $m + 1$ is represented by its $m + 1$ coefficients, i.e., a point in R^{m+1}. Let I be the cartesian product of I_i, $i = 0, 1, 2, \ldots, m$. I is a compact subset of R^{m+1}. Define

$$H = \begin{bmatrix} 1 & y_0 & y_0^2 & \cdots & y_0^m \\ 1 & y_1 & y_1^2 & \cdots & y_1^m \\ & & \cdots & & \\ 1 & y_m & y_m^2 & \cdots & y_m^m \end{bmatrix}$$

The matrix H is a mapping from the space R^{m+1} of coefficients into the space R^{m+1} of values.

H is a nonsingular Van der Monde matrix [13]. Thus the set $H^{-1}[I]$ is a compact subset of R^{m+1}. This set represents the set of all polynomials of degree less than $m + 1$ that satisfy the compact interval constraints I_i, $i = 0, 1, 2, \ldots, m$. Consider the map

$$G : H^{-1}[I] \times X \to R,$$

where

$$G(\alpha, x) = \sum_{i=1}^{m} i a_i x^{i-1}, \; x \in X,$$

and where $\alpha \in H^{-1}[I]$ is represented by $\alpha = (a_0, a_1, a_2, \ldots, a_m)$. The value $G(\alpha, x)$ is the derivative at x of the polynomial represented by α. Define

$$K = \max \{|G(\alpha, x)| : \alpha \in H^{-1}[I], \; x \in X\},$$

which exists since G is a continuous map with compact domain. The mean-value theorem establishes the assertion of the lemma.

In the next lemma we show how to construct an interval, say I, about $f(x)$, $x \in X$, such that if $g(x)$ is set equal to any value in I, then the event E occurs with a probability greater than P_s whenever x is the input.

Lemma 2. *Let* $x \in X$, *and let the input* x *be administered* T *times, resulting in the observations* $\Theta_j(x)$, $j = 1, 2, \ldots, T$. *Let* T *be an integer large enough so that*

$$\Theta_s(x) - \Theta_l(x) \geqslant 2\epsilon P_s,$$

where

$$\Theta_s(x) = \max \{\Theta_j(x) : j = 1, 2, \ldots, T\}$$

and

$$\Theta_l(x) = \min \{\Theta_j(x) : j = 1, 2, \ldots, T\}.$$

If $g(x)$ *is chosen so that*

$$\Theta_s(x) - \epsilon \leqslant g(x) \leqslant \Theta_l(x) + \epsilon,$$

then whenever the input x *is administered the probability of the event* E *occurring is greater than* P_s.

Proof. We want to show that

$$(1 - P_s)2\epsilon \geqslant |f(x) - g(x)|.$$

Since

$$\Theta_s(x) - \epsilon \leqslant f(x) \leqslant \Theta_l(x) + \epsilon,$$

it follows that

$$|\Theta_l(x) + \epsilon - \Theta_s(x) + \epsilon| \geqslant |f(x) - g(x)|.$$

The lemma follows since

$$(1 - P_s)2\epsilon \geqslant 2\epsilon - [\Theta_s(x) - \Theta_l(x)]$$

and

$$2\epsilon - [\Theta_s(x) - \Theta_l(x)] \geqslant 0.$$

Theorem 3. *Let* (X, f, Y, ϵ) *be a black box with* $f \in \mathcal{F}$. *Let the values* P_s *and* $P_I(n)$ *be given. Let* P *be a prior distribution over* $\mathcal{F}$. *Then a model* $g \in \mathcal{F}$ *can be constructed such that* g *is an adequate model of* f.

Proof (algorithm). Such a model is constructable by means of the following algorithm. This algorithm consists of an iterative procedure of generating a candidate polynomial and then of testing the candidate. The iterations are halted whenever a candidate passes the test.

Before beginning the iterations, we generate an initial candidate p_0 to be tested in the first iteration.

Let x_0^* be chosen randomly from X. Set $y_0 = x_0^*$. Throughout this proof, x and y are used to denote inputs. Using Lemma 2 with $x = y_0$, apply the input y_0 until the inequality

$$\Theta_s(y_0) - \Theta_l(y_0) \geqslant 2\epsilon P_s$$

is satisfied. Define

$$g(y_0) = \frac{1}{2}\,[\Theta_s(y_0) + \Theta_l(y_0)]\,.$$

Let p_0 be the constant polynomial with the value $g(y_0)$. Note that

$$\Theta_s(y_0) - \epsilon \leqslant p_0(y_0) \leqslant \Theta_l(y_0) + \epsilon\,.$$

Let $m = 0$, and proceed to the testing phase.

Testing. In general, m is a nonnegative integer. For $i = 0, 1, 2, \ldots, m$ the values $g(y_i)$ have been established for the inputs y_i. We have a candidate polynomial p_m of the kth degree, $0 \leqslant k \leqslant m$, such that

$$p_m(y_i) = g(y_i), \qquad i = 0, 1, 2, \ldots, m\,.$$

Let n satisfy Eq. (5) for the probability $P(f = p_m)$. For $j = 1, 2, \ldots, n$, choose randomly x_j^* from X, and observe the output $\Theta(x_j^*)$. If the constraint equations

$$|\Theta(x_j^*) - p_m(x_j^*)| \leqslant \epsilon, \qquad j = 1, 2, \ldots, n\,,$$

are met, then the polynomial p_m passes the test; set $g = p_m$, and we are done. If not, then let j_1 be the smallest integer in $\{1, 2, \ldots, n\}$ such that

$$|\Theta(x_{j_1}^*) - p_m(x_{j_1}^*)| > \epsilon\,.$$

(Of course, if this inequality holds for some $j_1 < n$, then $x_{j_1+1}^*$ is never selected nor applied as an input.) Proceed to the modeling phase.

Modeling. Let $y_{m+1} = x_{j_1}^*$. Using Lemma 2 with $x = y_{m+1}$, apply the input y_{m+1} until the inequality

$$\Theta_s(y_{m+1}) - \Theta_l(y_{m+1}) \geqslant 2\epsilon P_s$$

is satisfied. Define

$$g(y_{m+1}) = \frac{1}{2}\,[\Theta_s(y_{m+1}) + \Theta_l(y_{m+1})]\,.$$

Let ρ be the smallest integer in $\{k, k+1, \ldots, m+1\}$ such that there is a polynomial $\hat{p}_{m+1}$ of the ρth degree satisfying the constraint

$$\Theta_s(y_i) - \epsilon \leqslant \hat{p}_{m+1}(y_i) \leqslant \Theta_l(y_i) + \epsilon \tag{6}$$

for all $i = 0, 1, 2, \ldots, m + 1$. (Recall that $f(y_i)$ satisfies the same set of constraints.) Of all such polynomials satisfying inequality (6), let p_{m+1} be one that minimizes the functional

$$\sum_{i=0}^{m+1} \sup \{|p_{m+1}(y_i) - q_i| : q_i \in [\Theta_s(y_i) - \epsilon, \Theta_l(y_i) + \epsilon]\}. \tag{7}$$

We redefine

$$g(y_i) = p_{m+1}(y_i)$$

for all $i = 0, 1, 2, \ldots, m + 1$. Return to the testing phase.

Note that if $\rho = m + 1$, then

$$p_{m+1}(y_i) = \frac{1}{2} [\Theta_s(y_i) + \Theta_l(y_i)]$$

for all $i = 0, 1, 2, \ldots, m + 1$. This follows since there is a unique polynomial of the $(m + 1)$th degree passing through the midpoints of

$$[\Theta_s(y_i) - \epsilon, \Theta_l(y_i) + \epsilon], \quad i = 0, 1, \ldots, m + 1.$$

The midpoints of these intervals give a minimum value to the functional (7).

Since the polynomial p_{m+1} satisfies inequality (6) for all y_i, $i = 0, 1, 2, \ldots, m + 1$, it follows from Lemma 2 that the event E occurs with a probability greater than P_s whenever any of the inputs y_i, $i = 0, 1, 2, \ldots, m + 1$, are administered.

Let N be the degree of the polynomial f. Since the polynomial f satisfies the constraints of inequality (6) for all inputs y, it follows then from the definition of ρ that $\rho \leqslant N$. The modeling phase, therefore, never produces a polynomial candidate whose degree is greater than that of the black box. Each succeeding iteration produces a better candidate in the sense that more constraint equations are satisfied. Before each modeling phase, the old candidate p_m satisfies m constraint equations. After the modeling, the new candidate p_{m+1} satisfies an additional constraint equation.

We now establish that the above iterative procedure terminates with probability 1 as the number of iterations goes to infinity.

Let m increase without bound. Let Z denote the nonnegative integers. Let $\{y_i : i \in Z\}$ be the set of inputs generated in the previous procedure. The probability is 1 that this set is dense in X. For each member y_i the constraints

$$\Theta_s(y_i) - \epsilon \leqslant f(y_i) \leqslant \Theta_l(y_i) + \epsilon \tag{8}$$

are satisfied. Using Eq. (1), note that

$$\Theta_s(y_i) = f(y_i) + \max\{\eta_{ij} : j = 1, 2, \ldots, T_i\}$$

and

$$\Theta_l(y_i) = f(y_i) + \min\{\eta_{ij} : j = 1, 2, \ldots, T_i\},$$

where η_{ij} is the error in the observation obtained for the jth application of the input y_i.

Define

$$\alpha_i = \max\{\eta_{ij} : j = 1, 2, \ldots, T_i\}$$

and

$$\beta_i = \min\{\eta_{ij} : j = 1, 2, \ldots, T_i\}.$$

Thus

$$\Theta_l(y_i) + 2\epsilon - \Theta_s(y_i) = 2\epsilon + \beta_i - \alpha_i . \tag{9}$$

Let $\lambda > 0$. For $j = 1, 2, \ldots, N + 1$, there exists with probability 1 an integer $i(j)$ such that

$$2\epsilon + \beta_{i(j)} - \alpha_{i(j)} < \lambda .$$

Let

$$t_1 = \max\{i(j) : j = 1, 2, \ldots, N + 1\}.$$

It follows from Lemma 1 that there is an integer K such that for all $m \geqslant t_1$

$$|p_m(x_1) - p_m(x_2)| \leqslant K|x_1 - x_2|$$

and

$$|f(x_1) - f(x_2)| \leqslant K|x_1 - x_2| \tag{10}$$

for all $x_1, x_2 \in X$. Decompose X into a union of disjoint intervals X_k, $k = 1, 2, \ldots, J$, where J is some finite integer larger than $N + 1$, such that

$$K\,\mu(X_k) < \lambda, \tag{11}$$

where $\mu(X_k)$ is the length of the interval X_k. For each $k = 1, 2, \ldots, J$, there exists (with probability 1) an integer $i(k)$ such that $y_{i(k)} \in X_k$ and

$$2\epsilon + \beta_{i(k)} - \alpha_{i(k)} < \lambda . \tag{12}$$

Let

$$t_2 = \max \{i(k) : k = 1, 2, \ldots, J\}.$$

Define $t = \max \{t_1, t_2\}$. From inequalities (6), (8), (9), and (12) it follows that

$$|f(y_{i(k)}) - p_t(y_{i(k)})| < \lambda. \tag{13}$$

The inequalities (10), (11), and (13) together with the triangle inequality imply that

$$|f(x) - p_t(x)| < 3\lambda \tag{14}$$

for all $x \in X$. With p_t as a model of f, we have from inequality (14) that

$$P(E|x) > 1 - \frac{3\lambda}{2\epsilon}$$

for all $x \in X$. Thus

$$P(E|f \neq p_t) > 1 - \frac{3\lambda}{2\epsilon}.$$

Since $\lambda > 0$ was arbitrary, it follows that $P(E|f \neq p_t)$ converges to 1 as λ converges to zero. As a result, the probability of passing the testing phase goes to 1. Actually the polynomial p_{m+1} converges to f as m goes to infinity. This completes the proof of the theorem.

Consider the vector case where f is a polynomial function defined on $X \subset R^{m_i}$ and has range in R^{m_0}. For $k = 1, 2, \ldots, m_0$, let f_k denote the kth component of f, that is, $f = (f_1, f_2, \ldots, f_{m_0})$. Then f_k is a polynomial function from X into R. Let a vector $x \in X$ be denoted by its components $(x_1, x_2, \ldots, x_{m_i})$. Let $j \in \{1, 2, \ldots, m_i\}$. By varying the argument x_j and holding all other arguments of x fixed, the polynomial $f_k(x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_{m_i})$ can be estimated according to the previous algorithm. Consequently f_k can be estimated and therefore the identity f as well.

7. CONCLUSION

We have investigated the problem of identifying a black box whose input-output relationship f represents a static process having uniformly distributed errors in the observations of the output. We have assumed unlimited freedom in selecting test inputs from a compact subset of the input space. Identification of f is conducted by comparing the observations of the black box's output with that of a model. Bayes' formula is used to derive an equation for the *a posteriori* probability of identification as a function of *a priori* probability of identification, a probability of observing the output of the black box within error bounds, and a number of conducted tests (all tests being successfully passed).

We have defined a criterion for determining an adequate model of a black box. A model is defined to be adequate if it passes a series of n tests, where n is a given function of three quantities. Each test consists of administering randomly chosen inputs and observing the difference between the output of the black box and the output of the model.

The test is passed if the difference is less than the error bound of the noise. Sequences of distinct random inputs are shown to provide more useful knowledge about a model being adequate than are sequences of inputs in which an input appears more than once. The integer n is an increasing function of the probability of a successful test and the *a posteriori* probability of identification. That is, a more adequate model will satisfy more tests.

We have presented an identification algorithm for determining an adequate model of a black box that is representable as a polynomial function of a single variable. The algorithm consists of a procedure that iterates between a modeling phase and a testing phase. The procedure is halted whenever an adequate model is obtained. A discussion is given on generalizing the algorithm to handle vector-valued polynomials. This algorithm can be used to construct an adequate model for black boxes in which f is an essentially bounded Borel-measurable function, since polynomials approximate such functions within any degree of accuracy.

REFERENCES

1. L.A. Zadeh, "On the Identification Problem," *Trans. IRE* CT-3, 277 (1956).
2. L.A. Zadeh, "From Circuit Theory to System Theory," *Trans. IRE* **50**, 856 (1962).
3. P. Eykhoff, P.M. Van der Grinten, H. Kwakernaak, and B.P. Veltman, "Systems Modelling and Identification," *in Proceedings of the Third IFAC Congress, London, 20-25 June* (1966).
4. P. Eykhoff, "Process Parameter and State Estimation," *Automatica* 4, 205 (1968).
5. M. Cuenod and A.P. Sage, "Comparison of Some Methods Used for Process Identification," *Automatica* 4, 235 (1968).
6. A.V. Balakrishnan and V. Peterka, "Identification in Automatic Control Systems," *Automatica* 5, 817 (1969).
7. K.J. Åström and P. Eykhoff, "System Identification — a Survey," *Automatica* 7, 123 (1971).
8. P. Eykhoff, *Identification and System Parameter Estimation*, Part 1 and 2, North-Holland, N. Y. (1973).
9. P. Eykhoff, *System Identification: Parameter and State Estimation*, Wiley and Sons, N. Y. (1974).
10. H. Stalford and G. Leitmann, "On Representing a Black Box as a Dynamical System," *J. Math. Anal. Appl.* 38, 348 (1972).
11. E. Mark Gold, "System Identification via State Characterization," *Automatica* 8, 621 (1972).
12. H. Stalford and J. Kullback, "Identifying an Unknown Process by Using Randomly Chosen Inputs," *Information and Control* 23, 393 (1973).
13. O. Zariski and P. Samuel, *Commutative Algebra*, Vol. I, Van Nostrand, Princeton, 1958 p. 94.